AUTHORS’ NOTE: This paper is an extended version of a West Point senior capstone project paper by Sean Eyre, Randolph Rotte, and Theodore Taggart under the advisement of Paulo Shakarian, Christa Chewar, Anthony Johnson, and Patrick Roos. The original work by Eyre, Rotte, and Taggart was awarded “Best Computer Science Project,” “Best Network Science Project,” and the “Dr. George M. Krembs Innovation Award” at the 2012 West Point Project’s Day.

In 2010, the locals of the village of Gizab, Dykundi Province Afghanistan forcibly evicted the Taliban. Though not directly supported by U.S. ground forces, the villagers had received training and logistical support from U.S. Special Forces. Many hoped that the actions of the villagers could be replicated throughout the country. However, there are a limited number of U.S. Special Forces teams to conduct such missions. Can we identify a limited number of villages, such that if we provide them Special Forces support, the local revolts against the Taliban would be most likely to spread and stabilize at a large scale? In this paper, we take a network science approach to the problem. Using the tipping model from the social science literature, a network of villages created using tribal and spatial relationships, and a new software package called RASCAL (that the authors developed), we find that villages in Zabul have a significant influence on Helmand and Kandahar. In our results, villages in Zabul represented 40.00% of the influential villages for those three provinces – while representing only 26.25% of the total villages examined.

THE TIPPING MODEL AND THE HKZ VILLAGE RELATIONSHIP NETWORK

In the late 1970’s, Nobel-laureate Thomas Schelling introduced what is known as the “tipping model.” He originally used this model to describe how neighborhoods could become racially segregated. In his example, family of race A moves out of their household after a certain number of families of the race B move into neighboring homes. Upon their movement, they would likely be replaced by a family of race B as well. Hence, that house is “tipped” from race A to race B once a certain number of neighboring homes is occupied by families of race B. Schelling used this to illustrate why racial integration in neighborhoods is difficult. Around the same time as Schelling published his work, a notable social scientist named Mark Granovetter studied the idea of tipping on a social network. In Granovetter’s version of the model, an individual “tips” toward lineament (behavior) B after a certain number of his neighbors in the network adopted that behavior. We show this in the below illustration.
In the last decade, the tipping model has been studied in a variety of disciplines, including computer science (Kempe et al.), sociology (Watts), and economics (Jackson and Yariv). More recently, a compelling study by Damon Centola published in *Science* provides empirical evidence that tipping occurs in real populations. He showed that the tipping threshold increases the probability of an individual adopting a certain lineament or trait.

In this paper, we propose a new application of the tipping model. Like Schelling, we will use it to model the represent spatial relationships in a physical setting. However, we use a network structure like the work of Granovetter did to represent the relationships. We do this by creating a Village Relationship Network or VRN for the area we are considering. Given an area (in this paper we look at the three provinces of Helmand, Kandahar, and Zabul, Afghanistan), we create a network where each village is represented with a node. If two villages share a common tribe and are physically accessible to each other, then we draw an edge between them. This setup for the edges allows us to account for the difficult terrain of Afghanistan – if there is not a direct route between two villages, we assume there is no influence relationship. It also allows us to account for tribal influence. If two villages are inhabited by totally different tribes, then we assume that there is no influence relationship. We utilized road network and terrain data from the Afghanistan Information Management System (AIMS) and tribal data from the Naval Post Graduate School’s Program for Conflict and Culture Studies. The resulting network had 1,543 nodes and 1,087 edges.

**RASCAL: RESOURCE ALLOCATION FOR CULTURAL ADOPTION OF LINEAMENT**

In the last section we described the model we wish to use – the tipping model – as well as the dataset that we will use to represent village relationships in Helmand, Kandahar, and Zabul provinces (the VRN). Now the question is, suppose we have a certain number of Special Forces teams (we will designate this number as $K$) that we can place at any village in the three provinces. Can we position them in a manner that will cause the propagation of the anti-insurgent sentiment to spread to the maximum extent possible (assuming the tipping model)? Our software package RASCAL (Resource Allocation for Cultural
Adoption of Lineament). Given a VRN and limited number of Special Forces teams (designated by the number $K$), RASCAL assigns villages to the teams in a way that will cause the propagation of a behavior (via the tipping model) to spread to a large portion of the villages. We provide an overview of the algorithm in the appendix. In order to prepare this software for actual use, we also equipped it with a graphical user interface that allows for the user to select and modify the parameters for evaluating a chosen village network graph and have the results displayed in both visual and numerical forms. The software outputs both textual results as well as map results[6] for use in Geographical Information System (GIS) software such as Google Earth. One of the main goals of the user interface is to make the user comfortable with being able to differentiate between the separate algorithms that will analyze the village network graph. Therefore the interface has multiple ways that provide a basic understanding of the different algorithms. There are side panels in the interface that describe what each different algorithm does to answer questions by the user. There is also a help button that creates a help window that explains the basic functionality of the program and walks the user through using the program.

![Figure 2a: GUI Initialization Screen](image)
We conducted several experiments to see how RASCAL works compared to other methods. In our experiments we compared four different starting sets via the tipping model. These sets, all of size $K$, include a set of the top nodes when ranked by degree, a random set of nodes, and RASCAL. We tested sets ($K$) of size 5, 10, 15, 20, and 25. These were all run through the tipping mode with thresholds at 2, 3, 4, and 5 nodes (this threshold specifies the minimum number of neighbor nodes required in order for a node to adopt the new behavior). The below histogram displays how these three approaches performed (in terms of the total number of nodes that ultimately adopt the new behavior based on the tipping model).
Based on these results, RASCAL performed the best as it consistently led to the largest number of nodes in the network adopting the new behavior (under the tipping model). Random sets perform fairly consistently despite producing somewhat erratic results at times. Degree based sets perform well at low threshold values, but rapidly decline in effectiveness as the threshold increases—likely because high degree nodes are not necessarily always neighbors with other high degree nodes or even separate by a distance of one step. Thus, they do not form the cohesive units necessary to lead to cascades of individual adopting the new behavior. RASCAL handles higher thresholds much better, likely due to the location and locality of the initially selected nodes that start a series of adoptions of the new behavior.

Additionally, we analyzed the time required to complete each calculation. In order for RASCAL to actually be of operational use it must perform at a rate that does not detract from the operations tempo. Given that the algorithm is intended for resource planning and target development, we can spare more time than we could if the algorithm was intended for tactical use. In considering all villages of Helmand, Kandahar, and Zabul, using a standard laptop computer, RASCAL took under eight minutes of runtime. While this is a substantial amount of time, this is still operationally efficient to produce the level of effectiveness that the software is able to obtain.

<table>
<thead>
<tr>
<th>Set</th>
<th>Average Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>22.79795</td>
</tr>
<tr>
<td>RASCAL</td>
<td>443.2038</td>
</tr>
<tr>
<td>Random</td>
<td>22.8528</td>
</tr>
</tbody>
</table>

**RESULTS: THE INFLUENCE OF ZABUL**

As we experimented, we analyzed our data with the RASCAL because of its effectiveness and
theoretically reliable results. Using our algorithm we conducted tests with the same parameters previously tested, though this time with the intent of examining the real world context of the results. Namely, we tested target set sizes of 10, 15, 20, and 25 with tipping thresholds of 2, 3, 4, and 5. We began our analysis with the examination of a set of 15 villages at a threshold of 3 and discovered that Zabul is over-represented in the result – potentially indicating that this province may be “key terrain” with respect to Helmand and Kandahar.

Zabul makes up 26.25% of the full data set, but comprises 40.00% of RASCAL’s results. According to the Intelligence Knowledge Network, or IKN, at the Ft. Huachuca Culture Center, Afghans typically base their decisions on a number of factors, the most important of these being their religion, tribal ties and family. Every district in the province of Zabul has a Pashtun majority, which is not true for Helmand or Kandahar. Although it may seem counterintuitive for Zabul to have this much influence because of its homogeneous cultural makeup, it makes sense because both Helmand and Kandahar still have a Pashtun majority – which in turn may be more influenced by the Pashtun stronghold of Zabul. Utilizing the Worldwide Incident Tracking System Database, which logs all terrorist attacks worldwide, we calculated the attacks per village in each district. Helmand had 2.21, Kandahar 1.73, and Zabul 1.26. A possible explanation for the relatively low amount of attacks in Zabul could be that the Taliban already understood the importance of Zabul due to their knowledge of the area, and since they had previously taken steps to control that province, they did not need to conduct attacks in that province as much as Helmand or Kandahar. At one point, Taliban were so prolific in Zabul that they were able to collect taxes from the local populace as late as 2010. We also note that Taliban forces are able to infiltrate into the neighboring provinces after entering through Zabul, as they did in 2002-2003, although the mountainous terrain between Zabul and Pakistan makes travel between the two difficult. This rugged terrain also leads to a lack of official crossing points, making Zabul a prime location for the Taliban to re-enter Afghanistan after fleeing into Pakistan.
Figure 5: Results of a RASCAL run. This RASCAL run delivered a set of 15 villages, believed to be the most influential for changing behavior in the region. Six of these were in the east in Zabul while the other nine spread across Helmand and Kandahar.

DISCUSSION

While our results are interesting, we note that, due to the availability of useful, unclassified, open-source data, the village relationship network (VRN) used in this paper is incomplete in some respects. For instance, we only consider tribal relationships and road network accessibility among villages and ignore various other political, military, economic, social, information, and infrastructure (PMESII) characteristics that may link villages. The creation of a richer dataset with respect to these factors would likely enhance results – as well as introduce several other issues not present in this simple model. For instance, would we need to weigh an economic connection between two villages more than a political one – or vice-versa? Another important issue that would arise from such a dataset would be the presence of negative relationships. If two villages have a history of conflict, would the adoption of a certain behavior by one result in the other not adopting? Under what conditions can problems of this sort be overcome? The addition of information described in this section is an important direction for future work and the authors of this paper would also be interested if any members of the Small Wars Journal (SWJ) readership has access to such data – whether open-source or classified.[13]

Another aspect in which the dataset used in our experiments is incomplete would be to explore a dataset that includes more provinces than just Helmand, Kandahar, and Zabul. For instance, including areas such as the Uruzgan provinces or even parts of Pakistan may help provide more insight. Again, here we were limited by the availability of data.
Would adding additional types of relationships or provinces to the VRN significantly change our result of the inordinate amount of influence exerted by Zabul? Based on the tipping model and our algorithm, a larger and enhanced dataset might produce a different result – perhaps telling us that there are places more influential than what we found in our experiments. However, the additional data would only lead RASCAL to identify locations that would provide equal or better influence to the ones we found in our tests. So, while there could exist places with more influence of Helmand and Kandahar provinces than the villages we found (which included a significant portion in Zabul), what we found is still mathematically guaranteed to influence over 60% of the villages in Helmand, Kandahar, and Zabul by the tipping model. This number would still hold for a larger and enhanced dataset.

Another concern worth addressing is that the results of our experiments could be the incorrect because the tipping model itself is somehow flawed or does not accurately represent the information on the ground. We note that the tipping model, though simple, is well-established in the field of sociology has been used for over three decades. The maturity of the model has seen a wide variety of applications – including reasoning about the spread of phenomena in a geospatial setting. Hence, the use of this model is reasonable for our purposes. However, it would be possible to employ a more complex model of propagation, perhaps using computational techniques to learn various “rules” for how some behavior cascades through the villages. These rules would then comprise a richer model of how such a behavior cascades. This is a new and interesting area of scientific research that we are exploring. However, this work is still in its early stages.

Our lab currently is scheduled to begin a series of experiments where we will attempt to create a richer dataset using classified data and validate the results of RASCAL in a classified setting. Again, if there are members of the SWJ readership who are aware of such data, we would certainly be interested in further discussion.

CONCLUSIONS AND FUTURE WORK

In this paper, we have described a new piece of software we have utilized for identifying key villages for allocation of some resource (i.e. a Special Forces team) in order to encourage a cascade of a certain type of behavior. We applied this new software, called RASCAL, to the provinces of Helmand, Kandahar, and Zabul in Afghanistan and surprisingly found that Zabul was over-represented among the most important villages. Through its user-friendly design, RASCAL allows an analyst to apply the tipping model to village datasets, delivering results which can guide resource allocation decisions in Afghanistan and other counter-insurgency situations. Our continuing work further examines how various multidisciplinary and network science analytical techniques might similarly support counter-insurgency planning efforts, including examining alternative models for the diffusion of behaviors in a counter-insurgency environment.

ACKNOWLEDGEMENTS

Some of the authors of this paper are supported under OSD project F1AF262025G001 and ARO project 2GDATXR042. The authors are very thankful to these organizations for their support.

S.E., R.R., and T.T. would also like to thank LTC Fernando Maymi for his advice on this work throughout the course of the academic year.

We would also like to thank Small Wars Journal editor Peter Munson for his comments that helped improve this paper.

REFERENCES
Afghanistan Culture Smart Book IKN, Ft. Huachuca Culture Center. Available at: https://ikn.army.mil/.


APPENDIX: FURTHER DETAIL ON THE IMPLEMENTATION AND ALGORITHM OF RASCAL

The algorithm used in RASCAL combines the idea of shell decomposition, prevalent in mathematics[15] and physics[16] with a greedy covering algorithm – a technique from computer science. [17] Though we will provide technical details in a discipline-specific paper, we will provide the basic intuition here. First, the “core,” or the “nucleus” of the network is found by iteratively removing nodes (referred to as “decomposition”). Previous work has shown that, for social networks, nodes in the core tend to promote epidemic spreading of contagions. The below figure illustrates the different layers (known as “shells”) of a network that are examined en-route to finding the core.

The next step of the algorithm is to identify the $K$ nodes of the core of the network that are linked to as many other nodes in the core as possible. This is done by a greedy algorithm (see Hochbaum, 1985 for more on greedy algorithms). We implemented the procedure described above along with several other well-known methods for identifying important nodes in a network using a combination of Python and C# code to create RASCAL.


[6] These results are output using the standard Keyhole Markup Language (KML) format.


[13] Please contact Paulo Shakarian, paulo@shakarian.net if you have access to such data and are interested in provide it to us in order to explore it using RASCAL and/or related software.


About the Author

Paulo Shakarian et al.

This paper was composed by a team composed of students and faculty from the Department of Electrical Engineering and Computer Science, the Department of Mathematical Science, and the Network Science Center at the U.S. Military Academy in West Point, NY and the Department of Computer Science at University of Maryland, College Park, MD. Please direct all correspondence on this article to the corresponding author, Paulo Shakarian: paulo[at]shakarian.net.

Sean Eyre graduated from the U.S. Military Academy in May, 2012 with a Bachelor of Science in Computer Science.

Randolph Rotte graduated from the U.S. Military Academy in May, 2012 with a Bachelor of Science in Information Technology.

Theodore Taggart graduated from the U.S. Military Academy in May, 2012 with a Bachelor of Science in Computer Science.

Christa Chewar is a Lieutenant Colonel in the U.S. Army and holds a Ph.D. in Computer Science from Virginia Tech. Currently, she is an Academy Professor at the U.S. Military Academy.

Anthony Johnson is a Major in the U.S. Army and holds a Ph.D. in Mathematics. Currently, he is an Academy Professor at the U.S. Military Academy.

Patrick Roos holds a Master of Science in Computer Science from the University of Maryland, College Park where he currently is a Ph.D. Candidate in Computer Science.

Paulo Shakarian is a Major in the U.S. Army and holds a Ph.D. in Computer Science from the University of Maryland, College Park. Currently, he is an Assistant Professor at the U.S. Military Academy.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Military Academy, the Department of the Army, the Department of Defense, the United States Government, or any of the listed funding agencies.

Available online at: http://smallwarsjournal.com/jrnl/art/using-rascal-to-find-key-villages-in-afghanistan

Links: